A REVIEW OF COMPUTABLE GENERAL EQUILIBRIUM MODELS FOR TRANSPORT AND THEIR APPLICATIONS IN APPRAISAL

Edward N. Robson*
Research Centre for Integrated Transport Innovation (rCITI)
School of Civil and Environmental Engineering
UNSW Sydney
Sydney NSW 2052
Tel: +61 (2) 9385 5721 Fax: +61 (2) 9385 6139
Email: e.robson@unsw.edu.au

Kasun P. Wijayaratna
Research Centre for Integrated Transport Innovation (rCITI)
School of Civil and Environmental Engineering
UNSW Sydney
Sydney NSW 2052
Tel: +61 (2) 9385 5721 Fax: +61 (2) 9385 6139
Email: k.wijayaratna@unsw.edu.au

Vinayak V. Dixit
Research Centre for Integrated Transport Innovation (rCITI)
School of Civil and Environmental Engineering
UNSW Sydney
Sydney NSW 2052
Tel: +61 (2) 9385 5381 Fax: +61 (2) 9385 6139
Email: v.dixit@unsw.edu.au

*corresponding author

Date of first submission: 10 June 2017
Date of first revision: 27 May 2018
Accepted: 7 June 2018
ABSTRACT
In the transport planning process, decision makers require reliable and informative appraisals to facilitate comparisons and determine if a proposal is worthwhile to society. The cost–benefit analysis is the most common form of appraisal, where benefits are primarily measured from the change in consumer surplus in the transport market. However, these benefits will only reflect maximum social welfare if markets operate perfectly competitively and without any market failures. There may be significant uncaptured impacts, known as wider economic impacts, which agencies are beginning to incorporate in appraisals using ad-hoc methods. Computable general equilibrium (CGE) models are an increasingly popular method for assessing the economic impact of transport, including both direct and wider economic impacts, as they can determine the distribution of impacts among every market and agent in the economy by simulating the behaviour of households, firms and others from microeconomic first principles. Aside from their traditional role estimating changes in macroeconomic variables, CGE models can provide a measure of welfare that guarantees no double counting and accounts for nth order effects. This paper reviews the full range of CGE models that have been applied to transport issues and discusses their role in transport appraisal. CGE models for transport have been developed in urban, regional and environmental economics as well as other fields, and each field has applied its own theory, assumptions and practices to represent the relationships between transport and the economy relevant to the field. This paper also discusses the general role of CGE modelling in transport appraisal, as well as theoretical and practical concerns regarding CGE modelling practice.

KEYWORDS
Computable general equilibrium modelling; transport appraisal; cost–benefit analysis; wider economic impacts
1. INTRODUCTION

Planners and engineers worldwide strive to improve transport networks as cities grow and technology advances. Given ever-present constraints on funding, the appraisal of proposed projects is vital to provide a rational basis for decision making. Cost–benefit analysis (CBA), in which the economic impacts as well as some of the social and environmental impacts of projects are monetised, remains one of the most popular methods to assess and rank projects. In a CBA, each impact is assessed separately, taking care to reduce the risk of double-counting.

A key issue is that there are many interactions in the economy that are not captured in this process. For example, new infrastructure can stimulate economic growth, which in turn generates additional transport demand that may alter the benefits of the project and complicate its evaluation. These concerns have been recognised since the genesis of the CBA approach:

*If investment decisions are so large relatively to a given economy... that they are likely to alter the constellation of relative outputs and prices over the whole economy, the standard technique [of CBA] is likely to fail us, for nothing less than some sort of general equilibrium approach would suffice in such cases.* —Prest and Turvey (1965).

In other words, significant transport projects can impact demand and supply in other markets, and therefore the transport market should not be treated as independent from the rest of the economy. The effects of this treatment have been assumed to be inconsequential in CBAs until recent years, possibly resulting in incomplete and misleading analyses.

Agencies have begun to incorporate these uncaptured impacts, known as ‘wider economic impacts’ (WEIs), in CBAs over the past two decades to strengthen the justification for transport projects. In some cases, WEIs can rival traditional (direct) impacts in scale. Most WEIs are estimated with a number of ad-hoc models, which has led to differing assessment practices between jurisdictions and the risk of double-counting impacts.

One particular type of model that has the potential to unify the estimation of WEIs is the computable general equilibrium (CGE) model. A CGE model simulates an entire economy by representing the supply and demand of every market. The central mechanism is that both supply and demand in each market are functions of all prices across other markets in the economy, not just their own price, meaning changes in one market affect all others. Solving a CGE model involves searching for a set of prices that results in equilibrium in all markets simultaneously, i.e. ‘general equilibrium’.

Furthermore, CGE models applied to transport can provide a framework to assess both direct impacts and WEIs within a single model. GDP, prices and other economic measures can be extracted as the models are built from fundamental microeconomic behaviour. This enables agencies to prioritise across transport projects and facilitates comparisons with proposals for government expenditure in other sectors. Planners can also identify the distribution of impacts when agents and markets are spatially disaggregated, and can measure welfare directly from utility functions, rather than using the transport market as a proxy. However, there are questions about what role CGE models should play in appraisal. Data and computational requirements can be prohibitive, especially when spatial detail is necessary. The operation of CGE models also tends to be a ‘black box’ where model mechanics are hidden or difficult to understand. Thus far, only the Netherlands has guidelines detailing the use of CGE models in WEI appraisal (Wangsness et al., 2016).

This paper has two aims: (1) to review the full range of CGE models applied to transport issues in the literature, and (2) to synthesise the case for applying CGE models in transport appraisal. Section 2 summarises existing transport appraisal methods and issues,
including conventional CBA practices, their limitations and the valuation of WEIs within CBAs. Section 3 introduces the concepts underlying CGE modelling and Section 4 provides a comprehensive review of existing CGE models applied to transport. Section 5 explores how CGE models can be incorporated in appraisals, identifies theoretical and practical concerns regarding their application and outlines directions for future research. Section 6 concludes the paper. In contrast with previous reviews of CGE models in transport, this paper examines CGE modelling within the context of transport planning and covers all strands of economics that have developed CGE models for transport.

2. CURRENT PRACTICE IN TRANSPORT APPRAISAL

Transport appraisal involves the comprehensive and consistent quantification of the impacts of a proposed transport project or policy. It is critical as a decision tool within the broader context of transport development (shown in Figure 1) as it enables planners and engineers to refine solutions, compare proposals and assess the worthiness of a project. However, Mackie et al. (2014) explain that in practice, human judgement and its shortcomings are often a confounding factor in appraising proposals, and projects are often influenced by the political environment. It is therefore vital to have an efficient and robust appraisal methodology that is both useful to decision makers to select projects, and that can be relied upon by the community to provide a fair assessment of the outcomes of government spending. State-of-the-art practice concerning transport appraisal is discussed by Mackie et al. and Thomopoulos and Grant-Muller (2012).

![Figure 1: Conceptual framework for transport development](image)

Since the early- to mid-20th century, CBAs have been the dominant method of appraisal for transport planning. The concepts of CBAs originated in the 1800s with Dupuit’s (1844) study of how to set optimum tolls for a bridge. Following WWII, the modern CBA was developed by governments as they became increasingly responsible for providing infrastructure to support the population boom across the western world (Prest and Turvey, 1965).

CBAs are used in addition to other appraisal methods depending on the context of the project, informational requirements and available data. Examples of other appraisal methods

---

1 In this paper, Dupuit introduced the ideas underlying demand curves and the measurement of welfare from them.
include: (1) cost–effectiveness analysis, when the benefits of different options are expected to be similar or cannot be quantified, and (2) multi–criteria analysis (MCA), when it is not appropriate to combine impacts as in a CBA (Annema et al., 2015). The consensus suggests that both CBAs and MCAs provide valuable information for decision makers, but either tool used in isolation is susceptible to inconsistencies across the decision making process (Thomopoulos and Grant-Muller, 2012).

2.1 Cost–benefit analysis

The principles underlying a CBA are intuitive. For a transport project, all of the significant social, environmental and economic impacts of a transport project are first identified and monetised where possible. Future impacts are then discounted to a present value so that decision makers can use metrics such as net present value, benefit–cost ratio (BCR) and internal rate of return to prioritise projects and determine if an individual project is worthy of funding (Boardman et al., 2006). In doing so, CBAs assess social welfare improvement, rather than financial viability, as many of the benefits and costs of transport projects are not incurred by users. CBA methodologies are generally flexible as the range of impacts evaluated can be scaled to suit each project.

The impacts considered in a transport CBA span users, service providers and the community. User impacts include travel time savings (in-vehicle, waiting and access times) and their reliability, cost savings (operating and maintenance costs for private transport, and changes in fares for public transport), and mode attributes (comfort, safety and others). Service provider impacts include capital and operating expenses, and community impacts include air and noise pollution, urban separation, accident reduction etc.

While capital and operating expenses may be forecasted with engineering judgement and historical costings, user impacts are far less tangible. In CBAs, a transport user’s welfare is indirectly measured as the change in ‘consumer surplus’; the consumer surplus being the difference between what users are willing to pay for a transport service and what they actually pay. The concept of ‘price’ here refers to an index of all of the measurable attributes of a transport service, rather than just monetary costs. The index is constructed by converting each attribute into a monetary equivalent by using weights and conversion factors, as determined from revealed or stated preference studies, and then aggregating them into a ‘generalised cost’ of travel.

For a transport service, the demanded quantity of trips \( d(p) \) can be taken as a function of generalised cost (the ‘demand curve’). The area between the demand curve in two dimensions and generalised cost will then represent the total consumer surplus of the service. To illustrate, consider a transport project that lowers the generalised cost of a service from \( p_1 \) to \( p_2 \), as shown in Figure 2. As a result of the reduction in generalised cost, existing travellers will gain a benefit equal to the rectangle \( A \). The improvement will also induce additional demand from redirected or newly generated trips, which will increase the quantity of trips from \( q_1 \) to \( q_2 \). These new users will gain a benefit equal to the triangle \( B \). The sum of \( A \) and \( B \) then represents the change in consumer surplus from the reduction in generalised cost. In practice, knowing points \((p_1, q_1)\) and \((p_2, q_2)\) is enough to calculate the change in consumer surplus since the demand curve is assumed to be linear between the two points (the ‘rule of a half’). Changes in demand and generalised costs (including travel times) are estimated with external transport models that account for mode choice and congestion. Adjustments to the basic consumer surplus measurement are often required, particularly when the perceived

---

2 For example, if a user is found to value travel time savings at $20 per hour, a 30 minute travel time saving would be calculated as a $10 reduction in generalised costs.
generalised cost does not equal the social generalised cost due to taxes and subsidies (which may lead to a service being under or overused), demand has been diverted from other infrastructure, or there are changes in the use of upstream and downstream infrastructure (Transport and Infrastructure Council, 2016).

![Figure 2: Demand curve $d(p)$ and consumer surplus](image)

However, there are two well-known weaknesses of this methodology: (1) it assumes that all benefits can be measured from the transport system, and (2) it cannot describe the eventual distribution of benefits among economic agents (Bröcker et al., 2010). Until the 1990s, transport networks were treated separately from the broader economy in transport planning—a partial equilibrium perspective. Land-use distributions and economic parameters were assumed to remain constant in four-step demand models, producing fixed trip matrices to feed into appraisals (Mackie, 2010).

A general equilibrium perspective would instead suggest that a change in the transport market, for example a price reduction from $p_1$ to $p_2$, would not only affect the demand for transport $d(p)$, but also demand and supply in external markets. This, in turn, would induce second-order effects on the demand for transport and the equilibrium generalised price.

Land markets have a particularly strong connection with transport networks (Wegener and Fuerst, 2004). If a transport improvement provides a travel time saving to an individual, they may utilise it by increasing their travel demand or by moving to a preferred residential location that has become more accessible, whilst maintaining a similar daily travel time (Ahmed and Stopher, 2014). Transport demand curves would therefore shift, land values would change and the benefits of a transport improvement would transfer from travel time savings to other benefits as people adjust their lives to maintain their travel budget (Metz, 2008). Therefore, the measurement and reporting of benefits in a CBA as travel time savings—up to 80% in road projects according to Mackie et al. (2001)—may be misleading.

Proponents of the CBA approach claim that regardless of where they eventually accrue, all impacts to users will be captured within the consumer surplus metric. If markets operate perfectly, where prices equal marginal costs, then transport users will correctly consider the changes in external markets in their valuation of transport (Dodgson, 1973; Jara-Diaz, 1986; Vickerman, 2007a). The willingness to pay for a transport improvement would
therefore account for both changes in property prices as well as travel time savings (Sue Wing et al., 2007). However, this assumption is unlikely to be realistic as any source of market imperfection, such as monopolistic firms, taxes and excess demand/supply, would violate it. Even if it were realistic, externalities such as agglomeration would still not be reflected. Appraisals nowadays consider these types of impacts (WEIs) to be separate to user impacts.

The second issue with the conventional CBA methodology is that it can be difficult to determine the eventual distribution of impacts. When general equilibrium effects are ignored, transport models can estimate how much a region will benefit from travel time savings. This distribution is likely to change once markets adjust to the transport improvement as the benefits transfer away from travel time savings and into other markets. Even if there is perfect competition so that consumer surplus captures aggregate welfare as claimed by Jara-Diaz (1986) and others, conventional CBAs measure consumer surplus from market demand curves and thus do not identify how much a particular household will benefit. This appears to be a serious issue regarding the usefulness of CBAs since some studies indicate that governments prioritise projects on the basis of the distribution of benefits as much as their magnitude (Odeck, 1996; Eliasson and Lundberg, 2012; Eliasson et al., 2015).

2.2 Wider economic impacts
WEIs are a range of newly-recognised economic impacts in markets not associated with transport, which cannot be captured in the partial equilibrium measurement of consumer surplus through the transport market demand curve. These occur when the price of transport does not equal its social marginal cost due to market imperfections and technological externalities. If this difference exists, welfare may be gained (or lost) by the consumer as the economy transitions towards general equilibrium.

WEIs came to the forefront of transport economics during the 1990s as concerns grew about biases in existing appraisal practices and governments sought to justify infrastructure projects on the basis of economic growth. The seminal SACTRA (1999) report identified sources of WEIs and recommended amendments to appraisal practice, including the requirement of an ‘Economic Impact Report’ to supplement conventional CBAs. Over time, these recommendations have been implemented somewhat irregularly. The field is still under active development and the nature, relevance, and even existence of WEIs are contentious. Some BCRs for large projects, such as Crossrail in the UK, have shown significant improvement after including WEIs (Banister and Thurstain-Goodwin, 2011).

The WEIs most frequently recognised in a transport appraisal are (1) agglomeration externalities, (2) labour market effects, and (3) impacts in markets with imperfect competition (Wangsness et al., 2016). Agglomeration externalities refer to the relationship between the concentration of economic activity and productivity (Venables, 2007). When businesses locate near each other, there are increasing returns to scale through knowledge spillovers, better access to markets and sharing of facilities. Transport projects reduce the effective distance between businesses, thereby generating productivity benefits that spread throughout the urban area. There are also WEIs from the labour market due to the presence of taxes, imperfect information and imperfect competition. When people enter the workforce, change jobs or their hours of employment as a result of a transport project, they do so on the basis of their net wages. However, the benefit to society can be measured by gross wages, and thus the additional taxes they pay are an otherwise uncaptured benefit (Wangsness et al., 2016). Finally, firms that set prices above marginal costs are another potential source of WEIs as they have additional scope to improve efficiencies, resulting in welfare gains. While the WEIs mentioned so far are generally benefits, WEIs can also include detrimental impacts to social welfare (Kanemoto, 2013), though these are not often discussed.
In a study of appraisal guidelines across 23 countries, Wangness et al. (2016) found that 15 so far acknowledge the existence of WEIs. The authors found remarkable disparity between the guidelines they reviewed, but a large number appeared to take inspiration from the UK Department for Transport (2014) guidelines. Many sets of guidelines considered the evidence for WEIs to be less robust than for other impacts. As a result, only seven sets of guidelines specified WEIs as a component of CBAs, and most recommended that a BCR without WEIs be presented first. In these guidelines, WEIs are generally calculated with a set of equations based on transport and economic changes (e.g. employment densities and generalised costs of travel), using parameters from other studies (e.g. elasticity of productivity with respect to employment) and rules of thumb as other inputs. The lack of consistency has also led to concerns about double-counting impacts. There appears to be a need to find alternative models and methodologies that authorities find more reliable to evaluate the impact of transport on the economy.

2.3 Economic models for transport
Interest in the economic effects of transport predates the recognition of WEIs, and a number of models have been developed to simulate interactions between transport and the economy. In terms of appraisal, these models produce outputs that can be used in MCAs, economic impact reports and, since the advent of WEIs, equations to estimate economic uplift. The most basic level of modelling involves surveying firms about their attitudes towards transport and their likely responses to specific projects. It tends to be difficult to draw meaning from these surveys as firms are motivated to answer strategically (Oosterhaven and Knaap, 2003). Production function models use the factors of land, labour, capital and some measure of transport as inputs into an aggregate production function to estimate regional economic activity. Early versions used infrastructure stock to represent transport, but were not able to distinguish between productive and unproductive infrastructure. Later models used accessibility and economic potential indicators instead, but difficulties in establishing causal relationships remained (Wegener, 2011). The most recent production function models include additional factors within a structural equation setup to enable linkages with other models and account for economies of scale.

Larger modelling packages do not fall into distinct categories as they integrate multiple modelling techniques and vary significantly in structure. Regional scale models are frequently based on a multiregional input–output (IO) structure that simulates trade flows between markets. These models use coefficients to describe the relationship between outputs and inputs of industries by region, allowing the effects of changes in demand and transport costs to flow through to changes in supply. IO frameworks have been criticised for a lack of flexibility in these coefficients, as well as a lack of supply-side feedbacks in prices and resource constraints (Bachmann et al., 2014). Many of these drawbacks have been addressed in recent IO developments, such as the RUBMRO class of models. In addition, some of these modelling frameworks have trended towards simulating general equilibrium (Wegener, 2011).

Urban modelling requires detailed consideration of both transport and land-use markets as these constitute two of the most important factors of urban spatial development. Land-use transport interaction (LUTI) models link explicit transport and land-use models to simulate feedbacks between the two systems over both short- and long-term time scales (Wegener, 2004). These models are detailed, mature, and are empirically-based. Location decisions within LUTI frameworks can be modelled with spatial interaction location models (to predict origins and destinations of trips), bid-rent location models (to simulate bid rents and market rents) and utility-based location models (to simulate the utility of locations) (Wegener, 2011). However, LUTI models have been described as lacking a strong
microeconomic foundation and, like IO models, rely on parameters that are fixed with regard to prices. It can also be difficult to extract welfare for the purposes of a CBA or incorporate details of imperfect markets (Oosterhaven and Knaap, 2003).

In general, economic models for transport tend to be empirically detailed but are restricted in terms of market representation. Most do not produce a metric for welfare and have a limited ability to address the requirements of decision makers who are increasingly interested in capturing the full range of impacts to households, businesses and the economy at large, as well as their distribution. CGE models have been employed for transport projects and policies to bridge this gap, and are the focus for the remainder of this paper.

3. OVERVIEW OF CGE MODELLING

CGE models can provide a single framework to incorporate all interactions between transport and the economy as they are built from first principles in microeconomic consumer and production theory. Despite some opinions to the contrary, welfare and nearly any other economic metric can be extracted as all prices in the economy are simulated and the behaviours of agents are founded on utility and production functions (Forsyth, 2014). The cost of this economic detail is that CGE models cannot match CBAs in terms of spatial or transport network detail, although the gap is closing as computing power increases. Thus, CGE models in the literature have tended to focus on high level analyses of transport projects and policies, varying greatly in formulation, application and integration with CBAs.

3.1 Background of CGE modelling

Modern CGE modelling has evolved from two distinct branches of economics: (1) IO modelling and (2) general equilibrium (GE) theory (Thissen, 1998; Mitra-Kahn, 2008). Models from the IO branch were coined as ‘computable general equilibrium’ (CGE) or Johansen-type models, whereas models from GE theory were known as ‘applied general equilibrium’ (AGE) or Arrow–Debreu models. Both ‘CGE’ and ‘AGE’ are used interchangeably in the literature nowadays.

The IO approach to CGE modelling arose from the works of Leontief (1941), whose IO models of the US economy comprised a set of linear equations describing the quantity of inputs required by each sector in the economy for a unit of output. These IO coefficients were derived from matrix representations of inter-industry flows (IO tables). For a given final demand, the IO models would predict the required production activity of each sector. Johansen (1960) introduced a price mechanism on the IO framework that allowed a more detailed specification of the behaviour of economic agents. Instead of being fixed, the IO coefficients in Johansen’s model were a function of price, allowing firms, households and investors to substitute between sources of inputs and outputs according to utility and profit maximisation. Further development of Johansen’s model slowed during the 1960s and early 1970s (Dixon and Rimmer, 2010).

In this period, the foundations of the Walrasian GE theory approach were being developed separately. Arrow and Debreu (1954) formalised a mathematically-rigorous model of an economy with production and consumption from standard assumptions in neoclassical economics. Their model was proven to have an equilibrium solution and was used to claim insights about the market system in general (Geanakoplos, 1987). Scarf (1973) developed algorithms to compute Arrow–Debreu equilibria, which enabled a number of researchers to apply the model to practical problems.

However, interest in the IO approach was rekindled in the late 1970s as it was considered to be more flexible and could be computed more easily compared to the GE theory approach, despite being relatively empirical. By the mid-1980s, the IO approach led
by the highly influential ORANI model (Dixon et al., 1982) became dominant, but not before adopting some of the terms and concepts of the GE theory approach.

Modern CGE models draw from both sets of literature, but the legacy of the two different approaches is still evident today. Models with an IO heritage are built from the perspective of applying behavioural equations to IO/national accounts data. As such, these models have features that would not be seen in models from the GE theory heritage, such as the ability to fix ratios of demand in the national income identity\(^3\). Furthermore, in these models there are typically more variables defined than equations, in addition to parameters. This requires the modeller to specify some variables as exogenous, known as the model’s ‘closure’, such that the number of variables equals the number of equations. The chosen closure depends on the purpose of the modelling e.g. to simulate short-term or long-term equilibria. On the other hand, models with a GE theory heritage are more closely aligned with the theoretical Arrow–Debreu (1954) model. In these models, there is often no variability in model closure and less use of exogenous agents and parameters, which may be seen as less flexible but provides consistency to allow more properties to be proven about the model. It is often possible to identify the philosophy that had the greater influence on a particular model by the stream of literature cited, as well as the naming convention for model variables and the selection of software.

There are also studies in the literature that are described as ‘general equilibrium’, referring to the more general property of analysing more than one market simultaneously. Only those models which encompass all measurable production in the specified model region, as given in IO tables and national or regional accounts, are considered in this review.

### 3.2 Fundamentals of CGE modelling

A CGE model comprises a set of equilibrium equations representing commodity markets in an economy. Demand and supply in each market are functions of every price in the economy, not just the price in their own market. CGE models are solved by finding a set of prices and outputs that results in equilibrium in every market simultaneously. All CGE models simulate the behaviour of consumers and the production process through representative households and firms respectively. Households are endowed with primary factors (e.g. labour and capital) that are sold to firms for an income. Firms then transform these factor inputs, possibly with intermediate inputs from other firms, into commodities. Households and firms purchase these commodities to provide utility and to produce further outputs. Households are assumed to be utility-maximising in their behaviour, and firms are profit-maximising, using utility and production functions that are known to behave well. Solving the utility and profit maximisation problems analytically yields supply and demand functions that are then used to compile the equilibrium equations.

Models vary greatly in terms of the sets of commodities, households and firms that are represented. Firms, for example, are always disaggregated to some extent by sector, and households can be disaggregated by skill and demographics. In addition to pure production and consumption, CGE models from the IO approach will typically include representative governments funded by ad valorem taxes on sales and production, as well as representative investors and external markets. Shoven and Whalley (1992) provide a frequently-cited guide for constructing CGE models, but in large part from the GE theory approach. As a result, there is little mention of the use of IO tables for calibration, even though most CGE models nowadays rely on IO tables for calibration.

---

\(^3\) The national income identity is: \(\text{GDP} = \text{consumption} + \text{investment} + \text{government expenditure} + (\text{exports} − \text{imports})\).
Once models are specified, they are implemented in software to access solvers or are specially coded. The two most popular software packages are GEMPACK and GAMS. In GEMPACK (Harrison and Pearson, 1996), models are expressed and solved using linear approximations to expedite calculations. In GAMS, a model is specified as the constraints of a nonlinear programming problem or as a complementarity problem using either the standard syntax or the specialised MPSGE syntax for CGE modelling (Rutherford, 1999).

To calibrate a model, parameters are back-calculated from a set of benchmark data representing an economy at equilibrium. This process ensures that the model will replicate the benchmark data perfectly when run without any changes. IO tables, often in their extended form as social accounting matrices, are used as the benchmark data source. Elasticities of substitution are often required by the functional forms as well and are derived separately. A weakness of CGE modelling is that these elasticities of substitution, as well as other elasticities that may be required, are normally borrowed from other regions rather than being empirically estimated for the applicable model region.

Static models are operated by altering exogenous parameters, known as ‘shocking’ the model, to produce a counterfactual equilibrium for comparative static analysis. On the other hand, dynamic models trace the transition of an economy over time. This is commonly achieved by simulating a sequence of short-term static equilibria where long-term parameters (e.g. capital stock) are adjusted with external models between time periods, typically years but sometimes as short as quarters. Some dynamic models maintain relationships between these quantities and other model variables, such as VURM which incorporates relationships between capital stock and investment as well as net foreign liabilities and the current account balance (Adams et al., 2015). Other models such as ORANI-INT solve for all time periods at once, rather than recursively (Malakellis, 2000).

Economic agents and commodities can also be identified by their location to give models a spatial dimension. These have been referred to as spatial, multiregional or inter-regional CGE models, depending on the field. In CGE modelling, regional varieties of a commodity (e.g. apples from regions A and B) are most commonly assumed to be imperfect substitutes. This is known as the Armington (1969) assumption, where agents will demand a combination of regional varieties depending on the relative prices of the varieties. The Armington assumption is also convenient for calibration as it allows for cross-hauling (two-way trade flows) of commodities, as is commonly observed in regional datasets.

4. REVIEW OF CGE MODELLING FOR TRANSPORT
The role of transport in the economy is to enable spatially separated entities to physically interact for a cost. Transport is therefore an integral component of spatial models. Correspondingly, it is generally necessary for CGE models applied to transport issues to have a spatial dimension, unless the model is used to assess economy-wide policies. A number of interrelated disciplines, such as urban economics and regional science, have adopted the theory and tools of CGE modelling to analyse transport issues, and each have their own set of standard practices. Bröcker and Mercenier (2011) provide a brief tutorial on developing CGE models for transport and outline the theories from these fields.

This review explores the academic literature relating to CGE modelling of transport by providing a background of each field and then discussing common characteristics, notable models and their application to transport. The review starts at the smallest scale of spatial disaggregation (urban CGE models), moving towards larger scales (regional CGE models) and finally to non-spatial models. Some models have characteristics that span multiple fields—these have been placed in the section that best fits their application.

Relevant papers were first identified in Google Scholar and other databases matching terms similar to ‘CGE’ and ‘transport’ with over five citations. Papers were also sourced
from previous reviews of CGE modelling in transport, in particular from Bröcker and Mercenier (2011), as well as from citations within all examined papers. These papers were then categorised according to their research fields and citations, and predecessor models were identified to determine the key papers in each research field.

4.1 Urban CGE modelling

Urban CGE models have been developed within the field of urban economics for the study of economic issues in urban areas. The field covers both theoretical and applied literature in spatial analysis, housing, government and labour (Mills et al., 2000). Although formal study of the economic issues of space dates back to von Thünen’s (1826) model of bid-rent curves, much of the modern field draws from the Alonso–Mills–Muth (AMM) model (Alonso, 1964; Mills, 1967; Muth, 1969) of a monocentric city. The AMM model became popular as it predicted that density and land prices would decrease away from a city centre, both of which were common observations at the time. Linear programming models of cities also appeared around the same period (see Mills (1972a)) as well as LUTI models. As described earlier, LUTI models have remained the most widespread type of formally-applied urban analysis model to date.

Following advancements in the 1970s, AGE models were adopted in urban economics to simulate urban economies—early examples include Mills (1972b) and MacKinnon (1974). These theoretical models represented hypothetical monocentric cities with simplified relationships to understand the relative effects of policies on urban economies. Within the urban CGE framework, models of residential land use (Arnott and MacKinnon, 1977), labour demand (Sullivan, 1983), agglomeration and decentralised employment (Sullivan, 1986) were developed.

Later, a series of papers by Anas and other collaborators linked urban CGE modelling with transport modelling to examine the effects of congestion on urban form. Anas and Kim’s (1996) model relaxed the assumption of monocentricity to study the formation and stability of multiple city centres. In their model of a hypothetical linear city, land was allocated endogenously to production, housing and roads, and household and firm activity generated commuting, shopping and freight trips. Travel times were also modelled endogenously using a congestion function. They tested the impact of scale economies in shopping, where consumers prefer to shop at larger centres, and demonstrated that this could lead to multiple, somewhat stable equilibria. Anas and Xu (1999) extended the framework to study the effects of job dispersion and Anas and Rhee (2006) compared the welfare effects of congestion tolls versus urban boundaries to control sprawl. More recently, these models were adapted to analyse the effects of carbon charges (Tscharaktschiew and Hirte, 2010), public transport subsidies (Tscharaktschiew and Hirte, 2012) and the economic impact of speed limits (Nitzsche and Tscharaktschiew, 2013).

The urban CGE models discussed above were all theoretical or used synthesised data. A major challenge has been to convert these into an operational form. Anas (1982) proposed the embedding of discrete choice models, in particular logit models, for location decisions within the CGE framework. Even though discrete choice functional forms are less common in other CGE fields, they are consistent with the continuous demand functions prevalent in CGE modelling e.g. the constant elasticity of substitution function (Anderson et al., 1989).

Applied urban CGE models tend to focus on the markets and agents relevant to urban microeconomic simulation, such as land markets, and less often on representative governments or external markets. In the typical setup, households choose residential and job locations according to discrete choice models, and conditional on those locations, decide how much and where to consume in terms of shopping, housing and leisure. Producers can also be competitors in land markets and regions tend to be geographically small. Locations are sorted
by assuming uniform utility or profit across the urban area—an assumption borrowed from
the AMM model. Equilibrium in this sense means that not only are supply and demand equal,
but agents also have no incentive to change location. The transport network often influences
household behaviour in these models through the use of a time constraint. Households
maximise utility subject to not only a monetary constraint, but also a time constraint where
travel times are traded off against leisure and labour⁴.

Perhaps the most well-known applied urban CGE model is RELU–TRAN, developed
by Anas and Liu (2007). RELU–TRAN comprises two modules—a CGE module and a
transport module—which feed into each other and iterate until convergence is achieved in
both modules. In the basic structure of the model, employment and consumption of goods in
the CGE module generate shopping and commuting trips which provide inputs for the
transport module. The transport module then determines mode split and assigns trips to the
transport network, returning expected travel costs to the CGE module.

In RELU–TRAN, the CGE module comprises four economic agents: households
(consumers), firms, landlords and developers. Households, exogenous in number and
distinguished by skill group, are modelled in a two stage utility maximisation process. In the
lower level, given residential location, job location and preferred housing type, households
maximise utility over consumption of goods, leisure and housing size, subject to monetary
and time constraints. In the upper level, households jointly choose their optimal residential
location, job location and preferred housing type according to a logit model. Producers,
distinguished by region and industry, minimise costs according to a constant returns to scale
production function of labour, capital, buildings and intermediate inputs. In addition,
landlords control the supply of floor space according on profitability, and developers
construct and demolish buildings according to demand.

RELU–TRAN has been applied to the urban areas of Chicago, Paris and Los Angeles
to model issues such as fuel price increases, cordon tolling and job growth from rail
investment (Anas, 2013), and has been extended to model fuel consumption (Anas and
Hiramatsu, 2012).

Rutherford and van Nieuwkoop (2011), and van Nieuwkoop (2014) more extensively,
similarly simulated a transport network, but instead formulated the mode split and traffic
assignment problems together with the CGE model as a single mixed complementarity
problem. This allowed the model to be programmed in off-the-shelf optimisation software
such as AMPL or GAMS, making the model easy to understand, implement and modify.
Robson and Dixit (2016) also used a mixed complementarity format to develop an urban
CGE model suitable for transport infrastructure appraisal. In their model, discretionary trips
were generated in addition to shopping and commuting trips, and freight costs were modelled
as proportional to travel times.

Other urban CGE models have introduced additional disaggregate choice structures to
simulate behaviour outside the pure CGE framework. The field of computable urban
equilibrium (CUE) modelling in Japan, which branched from Anas (1982) in the late 1980s,
utilises these extensively. CUE models lie in between CGE and LUTI models in terms of
economic consistency (higher for CGE models) and empirical detail (higher for LUTI
models)—see Ueda et al. (2012) for a review. These models share many features with
traditional CGE models in that agents are profit and utility maximising, and models seek
equilibrium between supply and demand. Both building and land space are modelled
explicitly in a CUE model, but unlike a CGE model, Walras’ Law does not hold. Some

⁴ Horridge (1994) presented an early example of an urban model using elements of this framework. Sato and
Hino (2005) added detail to the modelling of production activity by representing the five business activities of
administration, production, retailing of goods, retailing of services and distribution separately.
models, for example Muto and Ito (2006), integrate aspects of both standard CGE and CUE models.

TRESIS–SGEM (Hensher et al., 2012; Truong and Hensher, 2012) is the combination of a discrete choice model for travel behaviour (TRESIS) within a CGE framework (SGEM). In TRESIS, location and transport choices are modelled with nested discrete choice structures, given economy-wide variables such as housing prices. The economy-wide variables are then equilibrated using the continuous demand structure of SGEM. The system was developed with the rationale that discrete choice models can be more detailed and are easier to calibrate for household decisions. The full model was later used to estimate agglomeration impacts for a rail project in Sydney.

There are also urban CGE models that have focused on capturing agglomeration economies in greater detail. Ahlfeldt et al.’s (2015) model of Berlin incorporated agglomeration forces not only in production, but also in the residential choices of workers in terms of accessibility to employment.

4.2 Regional CGE modelling

Regional CGE models represent discrete regions at a scale larger than urban regions, with the set of regions modelled often spanning an entire country. Most of these models belong to the interdisciplinary field of regional science, which encompasses theories in economics and social science to study how regions develop and interact. At the regional scale, urban issues such as land markets, household transport demand and congestion may be less significant and are often not represented. There may also be more emphasis on macroeconomic behaviours. In regional CGE models, it is common to assume that households are immobile and labour forces from different regions are imperfect substitutes.

Interregional trade flows were historically modelled with gravity equations. Isard (1951) was the first to use an IO framework for regional modelling by extending Leontief’s IO model into the spatial dimension. Isard’s model has remained influential ever since in the mathematical modelling of regional economies, including regional CGE modelling. In these models, interregional trade flows are facilitated by freight transport. This induces a price difference between where a commodity is produced (where the seller receives the free on board, or FOB price) and where it is used (where the buyer pays the cost, insurance and freight, or CIF price). Transport is therefore a margin commodity in regional CGE models, and the impact of freight on the economy is most often the focus of regional CGE studies.

A major differentiator between regional CGE models is in how regional varieties of a commodity are treated. One field of regional CGE modelling has emerged from the disaggregation of single-region national models into the spatial dimension. These adopt the Armington (1969) assumption of regional varieties being imperfect substitutes and are commonly known as ‘spatial CGE models’ (SCGE models). Regional CGE models have also emerged from the conversion of spatial price equilibrium (SPE) models from a partial equilibrium to a general equilibrium basis. SPE models assume that regional varieties are perfectly substitutable, and this has carried over somewhat to their corresponding CGE models. Finally, the field of new economic geography (NEG) was spawned in the early 1990s as an attempt to integrate theories from regional science and urban economics. NEG explains the formation of cities as a balance between increasing returns to scale and increasing transport costs from agglomeration. Regional CGE models based on NEG assume monopolistic competition in regional production sectors.

---

5 Isard’s model assumed constant production coefficients and price ratios, like other IO models of the time. While these assumptions restricted its ability to model changes in transport costs, Isard’s model was successful in that it could feasibly (in a computational sense) capture complex interregional linkages for the first time.
Regional CGE models for transport from the three streams of SCGE modelling, SPE modelling and NEG are explored below. Donaghy’s (2009) more general review of SCGE models identified a similar distinction between streams of literature. While any regional CGE model could be applied to transport, this review focuses on those that have previously been applied to transport issues and describes how it was achieved.

4.2.1 Spatial CGE modelling

The first CGE models of national economies lacked a spatial dimension. In these models, transport was treated as a tradeable material input, created with a production technology similar to other sectors. This assumption could be justified at the macroeconomic level of analysis which did not rely on a concept of space. As CGE models were extended into the spatial dimension during the 1980s, prices could be differentiated by region. At first, transport was either ignored or continued to be treated as a production sector, but it was now possible to relate transport costs to the physical movement of goods. Existing regional economic models for transport evaluation, such as multiregional IO models, also moved towards general equilibrium by simulating the optimising behaviour of consumers and producers (Liew and Liew (1991) for example).

Buckley (1992) proposed one of the first SCGE models to incorporate the costs of transport and wholesaling services explicitly through price margins. Regionally produced goods in Buckley’s model were transported to clearinghouses in other regions where they became available for local production, consumption or export. Each movement between regions incurred a transport cost specific to each origin–destination (OD) pair, which was added onto the regional price. Around this period, Wigle (1992) simulated tariffs in Canada with two CGE model formulations: one with regions and transport costs, and one without. Overall welfare was found to be similar between the two models, but regional disparities were significant. Both studies showed that demand and supply for transport services differed greatly when regional dynamics were considered, compared to the production sector method used previously.

As theory and computing power improved through the mid-1990s, it became feasible to disaggregate models to the extent required for representing transport infrastructure, in addition to traditional applications in policy evaluation. Bröcker (1998) designed an SCGE model that included a number of simplifying assumptions to make the model easy to implement. These assumptions improved the model’s computational efficiency and allowed it to be implemented with data typically available in developed countries. In Bröcker’s model, transport agents in each region and for each sector import commodities from other regions, which are then combined according to the Armington assumption to form a ‘pooled’ commodity available for local use (the ‘Moses–Chenery’ assumption, from Chenery (1953) and Moses (1955)). The transport costs themselves are modelled as a price mark-up, known as the ‘iceberg assumption’ (Samuelson, 1954), which in effect assumes that an amount of the transported good is used up (‘melts’) during transit in proportion to freight costs. The convenience of this assumption is that production in the transport sector does not have to be specified. Both the iceberg assumption and the Moses–Chenery assumption are still frequently used in SCGE modelling. Miyagi and Honbu’s (1995) model, and its later extension and application for expressway appraisal by Miyagi (2001), similarly applied the pooling concept, with the addition of modal split for freight, household migration and household demand for transport.

These prototype models enabled governments to develop SCGE models for planning purposes. PINGO (Ivanova et al., 2002; Vold and Jean-Hansen, 2007), a model for freight movements in Norway, is based on Bröcker’s (1998) model but includes an explicit sector to provide transport services rather than assume iceberg costs. Using demographic inputs and
freight OD matrices, the model predicts freight growth rates and other long-term effects of transport projects in Norway which can then be used to inform more localised models\(^6\).

SCGE modelling of transport began to spread globally throughout the 2000s, with many models differing in their methods of simulating the transport sector. Kim et al. (2004) developed a dynamic model of South Korea to evaluate the regional economic impact of four highway proposals. In Kim et al.’s model, accessibility was a factor in the production function for each sector in each region. This was calculated as the distance-weighted population surrounding each region (Vickerman et al., 1999) to approximate the difficulty of transport from other regions. Sundberg (2005) used a static formulation with iceberg transport costs to model the impacts of the Öresund Bridge between Sweden and Denmark. Dakila and Mizokami (2007) applied shocks to the transport industry in a conventional SCGE model of the Philippines to determine the best modes and regions for transport investment. Ando and Meng (2009) analysed regional transport firms in an SCGE model of China.

There has also been significant research in SCGE modelling from Japan in addition to their urban models, particularly for studying the impacts of earthquakes and other natural disasters. Koike et al. (2000) and Ueda et al. (2001) developed an SCGE model to calculate the damage of the 1995 Kobe great earthquake. This model focused on passenger trips rather than freight, adopting the concepts of household time constraints and shopping trips used in urban models. Business trips also formed a production input. RAEM-Light (Tavasszy et al., 2007; Koike et al., 2009) was later developed to assess the spatial impacts of road investment as well as disaster scenarios, and was applied in the Netherlands, Japan and Hungary (Koike et al., 2012). While most of the formulation of RAEM-Light is relatively standard for CGE models, a logit model is used to govern interregional trade choices in a parallel of the Armington assumption. Tatano and Tsuchiya (2008) simulated the short-run economic impacts of the 2004 Niigata-Chuetsu earthquake by holding labour and capital inputs fixed, and adjusting transport costs to reflect damaged infrastructure. Outside Japan, Zhang and Peeta (2011) combined a multilayer infrastructure network with an SCGE framework as a generalised method to capture the dependencies between infrastructure systems (transport, communication, power, energy and water) that become critical in extreme events.

The models described thus far were formulated with special consideration of the transport sector, but there exists a wide range of more general SCGE models that have been applied for transport analysis as well. Most are direct descendants of the IO/Johansen lineage, derived from CGE models created by the Centre of Policy Studies in Melbourne (e.g. ORANI, MONASH, TERM and MMRF) and GTAP (Hertel, 1997) and its databases. There tends to be more emphasis on macroeconomic impacts in these models, and accordingly they have a national or international scope with representative governments, investors and foreign sectors. Like in standard CGE models, transport is treated purely as an industry or commodity, with infrastructure improvements expressed as a technical change for the industry or a change in capital. Spatial disaggregation is achieved with both top-down formulations, where regional results are calculated from a high level simulation, and bottom-up formulations where the regions interact directly with each other.

The TERM model (Horridge, 2012) is one of the most regionally-detailed examples of this style of model. In TERM, the transport margin for a commodity is specified by both the source and destination of the commodity, as well as the region of use of the margin. The original TERM model was calibrated for Australia, but has since been adapted to models of China (SINOTERM), Indonesia (INDOTERM) and the US (USAGE-TERM).

---

\(^6\) Other examples include BROBISSE for road infrastructure in Denmark (Caspersen et al., 2000) and early models in the NEG literature described in Section 4.2.3.
In terms of transport applications, Schäfer and Jacoby (2005) coupled the EPPA model, an international CGE model of emissions built from the GTAP database, with a MARKAL (market allocation) model to assess climate policy while accounting for transport sector emissions. The MARKAL model is a dynamic linear optimisation model of the energy sector which represents the technological detail that cannot be accommodated in a CGE model. In the framework, transport demand, as generated from the EPPA model and a modal split model, informs the MARKAL model and influences substitution elasticities in the EPPA model. Karplus et al. (2010) used the EPPA model to analyse the environmental impact of introducing plug-in hybrid electric vehicles in the United States and Japan. Later, Karplus et al. (2013) applied an updated version of the EPPA model to predict global demand for passenger vehicle transport. New and used private vehicles were modelled as separate commodities to allow for fleet turnover and differing fuel consumption characteristics. Kishimoto et al. (2014) adapted parts of the EPPA model to develop C-REM, an SCGE model of China. C-REM was then used to investigate the impacts of an economy-wide carbon emissions policy with separate consideration of household vehicle transport. Verikios and Zhang (2015) analysed the effects of urban transport reform in Australia, including changes in governance, pricing and market structures, on household income groups using the multiregional MMRF model to simulate region-specific changes.

4.2.2 Spatial price equilibrium modelling
Spatial price equilibrium (SPE) models predict the production and flow of goods between regions in an economy to meet consumer demand, while accounting for transport costs. From the founding work of Samuelson (1952) and its definitive treatment by Takayama and Judge (1971) until the rise of CGE models, SPE models were the primary method for simulating the interaction of transport and the economy. Techniques from SPE modelling continue to be incorporated in CGE models today. Furthermore, SPE models have been prominent within transport engineering for freight flow estimation. The mathematical programming formulations of Beckmann et al. (1956) are highly influential in both SPE and traffic assignment literature.

An SPE model comprises a set of producers and consumers, with each producer and consumer representing a discrete region. Producer and consumer behaviour is expressed through inverse supply and demand functions in each region respectively. Commodities flow from producers to consumers and can be transferred between regions via ‘shippers’, but incur transport costs which may be fixed or a function of flows. This is normally formulated as a nonlinear programming problem, where solving the model involves finding a set of equilibrium prices and flows such that all supply and demand is satisfied. At equilibrium, commodities will flow from one region to another if the price in the receiving region equals the price from the supplying region plus the transport cost between the two regions.

This description relates to the SPE problem alone. However, the flow of commodities in an SPE model also generates freight trips. Freight network equilibrium models parallel traffic assignment in predicting how freight trips flow through a transport network via carriers. Harker (1987) integrated trip generation and distribution from an SPE model with mode choice and assignment from a network equilibrium model to create one holistic model for intercity freight flows.

SPE models in general are known to be mathematically robust. On the other hand, from the assumption that commodities are homogeneous, they cannot account for cross-hauling between regions; commodities from different regions are perfect substitutes and consumers will choose the source with the lowest delivered price. Due to the use of supply functions, which themselves are difficult to estimate, market interdependencies in the production process are not modelled unlike in CGE models.
Roson (1993) proposed replacing the SPE component of the freight flow framework with Walrasian (AGE-style) economic equilibrium. In Roson’s study, formulations of both Walrasian equilibrium and Wardropian equilibrium (‘user equilibrium traffic assignment’ in transport engineering) were shown to have the same analytical structure. This suggested that the two could be integrated to form a single freight model that accounts for market interdependencies, but a number of operational difficulties regarding data and the solution methodology were raised. Roson (1996) went on to develop the MITER model for freight flows in Italy which integrated a CGE model, dispersed SPE model and freight network equilibrium model. In MITER, trade flows and regional demand are assigned to the transport network, from which transport costs are used to update prices and then quantities in the model. The three models then iterate until equilibrium is reached in prices, quantities and the transport network. Friesz et al. (1998) later formalised the integrated CGE and Wardropian equilibrium problem in both nonlinear complementarity and variational inequality formulations. Friesz et al.’s study also described how imperfect competition among freight carriers could be incorporated.

In the above studies, the SPE model block of the freight flow framework was replaced entirely with a CGE model. Other authors have proposed formulations that incorporate both SPE and CGE behaviours. Elbers (1996) argued that while commodities which experience cross-hauling are appropriately modelled using the Armington assumption as they constitute distinguishable varieties, some commodities (primary commodities for example) are not distinguished by their origin i.e. are perfectly substitutable. For these commodities, setting high substitution elasticities could lead to difficulties in solving the model, and would be better represented as an SPE problem. Elbers also discussed issues relating to how transport services are produced, and how the combined CGE and SPE model could be solved. Lofgren and Robinson (2002) went further by assuming all commodities follow the SPE paradigm of perfect substitutability and one-way trade flows, but within a CGE framework. In their stylised model of a developing economy, an urban region acts as a hub between rural regions and the rest of the world. Households and producers are utility and profit maximisers respectively, and due to perfect substitutability, regions can shift between importing and producing/exporting particular commodities.

4.2.3 New economic geography modelling
Since the foundation of regional science in the 1950s, ideas such as central place theory and base-multiplier analysis were developed to describe how cities form, but had never formed a fully cohesive framework. The field of new economic geography (NEG) (Krugman, 1991; Fujita et al., 1999) attempts to reconcile theories from regional science and urban economics by explaining that cities form as a balance between increasing returns to scale (drawing industries together) and increasing transport costs. Mathematical modelling is central to NEG practice. To make them tractable, NEG models incorporate the Dixit–Stiglitz (1977) form of monopolistic competition as well as the iceberg assumption for transport (described in Section 4.2.1). Aspects of NEG analysis require a general equilibrium framework and many NEG-based CGE models have been applied for regional transport appraisal.

At the core of NEG is the assumption that agglomeration is driven by industries that are imperfectly competitive and experience increasing returns to scale. In contrast, other fields of CGE modelling assume perfectly competitive and constant returns to scale industries, as is standard in neoclassical economics. The NEG assumptions lead to multiple equilibria, and therefore NEG-based CGE models must be analysed differently since the equilibrium reached may depend on the prior state of the system (path dependence). This effect was previously observed in relation to Anas and Kim’s (1996) study on scale economies in shopping, described in Section 4.1. The dynamics of the spatial economy are
described in terms of ‘forces’ (or feedbacks), which may entrench an equilibrium state or shift it towards a different one.

In addition to Dixit–Stiglitz monopolistic competition, agglomeration externalities have been simulated by applying agglomeration factors to production functions. With this method, a firm’s productivity will increase according to the size of its region, but firms do not perceive these externalities in their decisions and will behave competitively.

The concepts of NEG were formalised in a general equilibrium context by Venables (1996) to examine (1) how agglomeration is affected by economic integration (one aspect being the cost of transport between regions), and (2) how important an industrial base is for a region. In Venables’ analytical model, both low and high transport costs were found to support a dispersed spatial equilibrium, whereas intermediate levels of transport costs would support a concentrated equilibrium with agglomeration in particular regions. Kilkenny (1998) supported this conclusion, demonstrating using an NEG-based CGE model that a reduction in overall transport costs could promote rural development. In relation to appraisal, Nordman (1998) compared welfare for a link improvement in terms of total equivalent variations with the change in consumer surplus from the transport market, using a CGE model based on Hussain (1996) and Hussain and Westin (1998). The presence of increasing returns to scale resulted in the change in consumer surplus measure, which is comparable to traditional CBA, underestimating total benefits to the economy.

Mun (1997) combined aspects of NEG, urban economics and SPE in a model to determine how pre-specified cities would be distributed in terms of size. Mun’s model permitted free migration with a fixed housing supply in each city, and households were sorted according to equal utility. Economies of scale were simulated with the agglomeration factor method.

Venables and Gasiorek (1999) later developed a CGE model using NEG theories to analyse the supply-side effects of road projects funded by the European Commission. Their model used a standard form of household and government consumption, but some industries were monopolistically competitive through setting a constant price mark-up over marginal costs. Projects were simulated for four stages of economic adjustment: (1) direct (interregional trade flows held constant), (2) short run (trade flows adjusted, but locations held constant), (3) medium run (number of firms adjusted, but workers held constant), and (4) long run (all adjustments allowed).

Bröcker (1995) developed a similarly structured CGE model to Venables (1996) with monopolistic competition and iceberg transport costs, but in an applied context with a greater focus on how the model could be calibrated and solved. The paper also demonstrated that trade flows in the model would follow a gravity law form, as is commonly assumed in other spatial economic models.

The models from Bröcker (1995) and Bröcker (1998) (discussed in Section 4.2.1) formed the basis of CGEurope, a family of CGE models used to assess regional transport links in Europe (Bröcker et al., 2001, 2004). Two versions of CGEurope were eventually developed for the IASON project, which aimed to create a unified assessment framework for proposed transport projects and policies in the European Union. Unlike conventional CBA, this framework would determine all regional welfare changes as well as their distribution. CGEuropeI comprised a high number of regions but only two sectors—one for local goods in each region under perfect competition and constant returns to scale, and one for tradable goods under monopolistic competition and increasing returns to scale. Iceberg transport costs were derived from a logit model of mode choice for both freight and business travel. On the other hand, CGEuropeII was designed to handle an arbitrary number of sectors, but at the cost of fewer regions. Rather than pure perfect or monopolistic competition, sectors could be calibrated with a market form parameter to allow a continuum between the two states. The
household utility function also included a component of private passenger travel, and iceberg costs were used for freight. Both models featured a single primary factor with no public sector.

A model similar to CGEuropeI was later used by Bröcker et al. (2010) to investigate the spatial distribution of impacts generated by the TEN-T transport network projects, an initiative of the European Union to improve economic and social cohesion between constituent regions and countries. For these projects, the spatial distribution of impacts was a particular concern as it would determine whether European Union involvement would be justified (when benefits were predicted to spill over into jurisdictions not financing the project) and what role they should play. Like with all regional CGE models except those based on the SPE literature, the lack of an integrated transport network model to measure endogenous congestion was noted as a limitation. Korzhenevych (2010) converted this into a dynamic model and incorporated a wage curve (a negative relationship between wage and unemployment rates), which was later published in Bröcker and Korzhenevych (2013). In the dynamic model, formulated in continuous time with perfect foresight, households maximise utility over time subject to a flow budget constraint and firms maximise present values. This model stands in contrast with typical dynamic models which recursively solve static CGE equilibria, where intertemporal decisions are not optimising.

RAEM is a family of NEG-based CGE models of the Netherlands developed specifically for transport project appraisal. They may also be the only CGE models consistently applied for WEI estimation. Derivatives of RAEM have been developed for other countries, including ISEEM for Belgium (Heyndrickx et al., 2009), TIGER for the Benelux countries (Heyndrickx et al., 2011) and RAEM-Europe. The first version of RAEM (Knaap and Oosterhaven, 2011) was developed in the early 2000s to assess the benefits of a high speed rail link between the west and the north of the Netherlands. A general equilibrium perspective was required as the link aimed to relieve pressure in the dense west by enabling jobs and residents to relocate to the north. Like other NEG models, industries were monopolistically competitive, but a proportion of output in each industry could only be consumed locally. Transport was modelled with a modified iceberg assumption where both freight distances and passenger travel times influenced the transport markup. However, in contrast to other models, wages were set as equal across all regions in approximation to the centralised wage conditions of the Netherlands. Migration could only be simulated in a separate model (Elhorst and Oosterhaven, 2006).

The experience of developing RAEM-1 highlighted a serious problem with the iceberg assumption in spatial models. Since the iceberg assumption implies that each sector produces its own transport service using its own production function, a reduction in transport costs (from a new transport link, for example) could lead to a reduction in that sector’s production, whereas in reality an increase would be expected (Tavasszy et al., 2011). As a result, RAEM-2 (Thissen, 2005) modelled transport with its own production function. RAEM-2 also incorporated search theory for labour markets, where firms would set vacancies and the unemployed would find jobs with a certain probability. Pilegaard and Fosgerau (2008) investigated the implications of modelling labour search theory in transport appraisal and found that it would lead to additional benefits due to the mismatch between wages and the marginal product of labour. Migration was simulated in RAEM-2 by setting utility as equal across the country, much like in the urban models.

RAEM-3 (Ivanova et al., 2007) resembled more conventional CGE models with the introduction of international trade and sector-specific capital. The model was converted from static to recursively dynamic through savings, capital accumulation and technological progress. Trip generation was also extended to include passenger trips other than commuting
and shopping, RAEM continues to be influential and its features have been proposed in an extension of PINGO, a model for Norway described in Section 4.2.1 (Hansen, 2010).

Fan et al. (2000) developed a model that has been used as the foundation of REMI, a dynamic macroeconomic model with a CGE-like core widely used by transport authorities in the US and elsewhere. In Fan et al.’s model, firms experience increasing returns to scale and are monopolistically competitive, but land is consumed by households and is a factor in production. By constraining land, the model replicated effects known to urban economics such as bid-rent and density gradients. Solving the model involves transitioning variables towards equilibrium (‘evolution’) according to ‘laws of motion’, and during the evolution process, regional industries can appear and disappear. In addition to the starting configuration, the speed of transition was shown to influence which equilibrium is ultimately reached.

There have also been models which incorporate elements of NEG within a more standard SCGE framework. Haddad and Hewings (2005) extended an existing CGE model of Brazil with increasing returns to scale and non-iceberg (i.e. explicit) transport margins. This model was extended again in Haddad et al. (2010) to incorporate node costs for a study of port capacities, and was integrated within a wider modelling framework to assess the economic contribution of the São Paulo subway system (Haddad et al., 2015).

4.3 Congestion and externality modelling

From the 1990s, a stream of research at the intersection of regional science, urban economics, transport economics and environmental economics has applied CGE models to analyse the externalities of transport, particularly congestion and pollution, and policies to address them. The general equilibrium approach is appropriate for the modelling of externalities as it can allow for linkages between economic agents that do not interact directly. These models represent the transport market in significant detail, often incorporating endogenous congestion and mode choice. Congestion (and analogously for other externalities) can be simulated either by incorporating a congestion index impacting utility and production, or through the use of a household time constraint (Sue Wing et al., 2007), in conjunction with some function relating transport demand to capacity e.g. a link congestion function. As a result of this complexity, most models in this category lack a spatial dimension, and hence only aggregate effects on the economy can be determined.

Congestion pricing, one of the most common policies to be analysed by these models, has long been proposed by transport economists as a mechanism to address the congestion externality of road transport—that is, due to the limited space of roads and their public nature, each additional vehicle on a road incurs a cost to other users that is not borne by that additional vehicle. In a first-best setting, road users would simply pay for the marginal social cost of using the road (a Pigouvian tax), but economies are distorted in reality by taxes and imperfect competition. Thus, researchers have adopted the CGE approach to determine whether congestion pricing is effective in a second-best setting and whether it provides an opportunity to reduce other distortions—the so-called ‘double dividend’ (Munk, 2003).

Building on a series of models of externalities from environmental economics, Mayeres and Proost (1997) used the congestion index approach in developing a CGE model to assess the tax system of an economy with congestion. In their model, an economy-wide index of congestion is a function of consumer and producer use of externality-generating goods (i.e. cars and trucks) as well as investment in road space. The congestion index provides a direct disutility to consumers and reduces the productivity of firms. The model was later generalised and extended in Mayeres and Proost (2001) to consider additional tax instruments.
Congestion, accident and pollution externalities were then simulated together by Mayeres (2000) to test the efficiency of peak road pricing, fuel taxes and public transport subsidies in the context of lump sum transfers and labour taxes. This model used the household time constraint approach of simulating congestion, similar to the urban models in Section 4.1, and represented household behaviour in more detail than in any predecessor model. Using the DeSerpa (1971) model of time, households maximised a 10-level utility function comprising consumption, leisure and the time spent on consuming transport goods, i.e. travel time, subject to a monetary constraint and time allocation constraints. Additional and separable components of utility also included government spending on public goods and negative contributions from air pollution and accidents. Congestion was reflected in the model as an increase in travel times, which were valued endogenously according to wages and preferences over travel. These travel times were modelled using a link congestion function representing the road network. The results indicated that peak road pricing would be the most efficient mechanism overall, but fuel taxes would be more effective for reducing air pollution and accidents specifically. In an analysis of marginal social cost pricing versus average cost pricing, Mayeres and Proost (2004) extended the model to include road damage and increasing returns to scale for public transport industries.

Parry and Bento (2001) examined in detail how revenue from congestion pricing should be spent. Like Mayeres’ (2000) model, households maximised utility subject to time and monetary constraints, but transport demand was separable from consumption and leisure. Travel times were also a function of trips, with congestion charges levied per trip. Their CGE model suggested that if the revenue is redistributed as transfer payments, as is commonly assumed in other models, labour supply could be discouraged, leading to a welfare loss outweighing the gain from reduced congestion. On the other hand, an equivalent reduction in labour taxes could lead to a significant welfare gain from the improved efficiency of the labour market. Parry and Bento (2002) introduced suboptimal public transport pricing, congestion on competing routes and fuel taxes into the previous analysis. This model also incorporated accident and pollution externalities into household utility and modelled travel times using a link congestion function.

De Borger and Van Dender (2003) and Van Dender (2003) applied similar models to the Parry and Bento (2001) model but with the addition of household demand for leisure trip commodities. In the former, congestion pricing was examined for its effect on values of time, whilst the latter explored the welfare cost of charging both commuting and leisure trips the same congestion price. Munk (2005) continued this line of research by deriving optimal transport policies, and noted that transport and consumption should not be treated as separable in analyses of transport policies since they influence each other, unlike other environmental externalities.

To extend the Swedish EMEC model, Berg (2007) designed a detailed household utility function of travel based on empirical literature of mode choice. Transport products in Berg’s model were distinguished by trip purpose, mode and length, which permitted different behaviours for each product, and households were disaggregated by income and density of their residential area. Similar to previous models, commuting was complementary to labour supply, but congestion was not endogenous to the model.

Prior to RAEM-2 as described in Section 4.2.3, Pilegaard (2003) developed a CGE model incorporating labour search theory and congestion to analyse commuting subsidies, wage taxes, transport taxes and subsidies to firms for vacancies. Vandyck and Rutherford (2014) later integrated aspects from NEG in a study of the efficiency and equity of road pricing in Belgium with congestion, agglomeration and unemployment. Similar to other NEG studies, Vandyck and Rutherford argued that congestion pricing should be lowered when there are agglomeration externalities and other inefficient taxes. The model comprised rural
regions as well as an urban region, and commuting was modelled with the time constraint method. Salaried workers were mobile, but workers earning an hourly wage experienced a wage curve and were immobile. The study demonstrated that commuters themselves could experience a welfare gain from congestion pricing if the revenue is distributed appropriately.

Other models have gone further in representing the production of transport services and their interaction with industry. Conrad (1997) developed a model to determine optimal levels of transport infrastructure and investment in the context of congestion and taxes. Production in Conrad’s model comprised capital, labour, energy, material and transport, where the transport input itself was a combination of transport services and capital. Transport capital was a function of the stock of trucks and the availability of infrastructure. Productivity of transport was diminished by a congestion index equal to the ratio of transport capital to a baseline value. In a multi-step analysis, Conrad gradually introduced constraints and mechanisms into the model to calculate the optimum level of infrastructure as funded by a fuel tax, which itself incurred a dead-weight loss. Conrad and Heng (2002) used a similar model to examine whether increased fuel taxes to finance the reduction of road bottlenecks could be covered by decreased congestion costs.

Sue Wing et al. (2007) combined a model of freight services with the household time constraint approach of congestion. In Sue Wing’s model, inter-industry transactions required transport services for freight. However, in congested conditions, transport firms would service fewer trips due to their capital stock (vehicles) being spread more thinly when travel times rose. Trips were allocated to links, introducing a spatial element, though it is unclear what the links represented. Households also demanded both commuting and shopping trips to supply labour and consume goods.

4.4 Other non-spatial CGE modelling

The following is a brief description of other non-spatial CGE models that have been applied to transport issues.

Steininger (2002) used a conventional CGE model with foreign trade to assess the impacts of a tax per tonne-kilometre of road freight in Austria. Johnsson (2003) modified this model to simulate road wear and kilometre charges in a thesis on transport tax policies. Steininger et al. (2007) later analysed a car road pricing policy applied nationwide in Austria, where the road price entered household utility as a variable cost of private household transport.

Siegesmund et al. (2008) evaluated maritime infrastructure investment in Texas using the USAGE-ITC model, adapted from the MONASH model of Australia, where the investment represented a technical change to the water transportation margin commodity. Dixon et al. (2017) extended USAGE to accept inputs from the highway model HERS, forming the USAGE-Hwy model. This model introduced highways/bridges, street repairs, private road transport, vacation transport, commuter transport and household car repairs into the USAGE framework, and the behaviour of transport industries already in the model were adjusted. USAGE-Hwy also incorporated a phantom tax to model travel time savings. This allowed the model to accept changes in travel times and costs from highway investments as simulated by HERS.

Based on the GTAP database, Chen and Haynes (2013) examined the role of public transport capital stock in the US and how different modes use public capital as an input. This modelling structure was also used to investigate high speed rail infrastructure in China (Chen et al., 2016).
4.5 Review summary
Each of the interrelated fields of CGE modelling explored above have applied their own theory, assumptions and practices to represent the relationships between transport and the economy relevant to their field.

Urban CGE models emphasise household transport behaviours and their impacts at the urban scale, including land markets and congestion through simulating commuting and shopping trips. Other characteristics of urban CGE modelling include (1) the use of discrete choice functional forms to model individual behaviour, (2) less emphasis on freight than in regional models as data at the urban scale is often near the level of individual transactions, and thus can be difficult to obtain, and (3) less emphasis or even no representation of macroeconomic agents such as governments, investors and external markets.

Regional CGE modelling commonly focuses on the relationship between freight costs and spatial development. In regional CGE models, transport is treated as a cost of interregional trade, either modelled explicitly as a margin commodity produced by a transport industry, or assumed to be produced by the industry of the transported good (the iceberg assumption). Three branches of regional CGE modelling have been identified: (1) SCGE models derived from the disaggregation of single-region CGE models, (2) CGE models derived from SPE models and (3) CGE models derived from NEG theory. Typical elements of SCGE models include perfect competition, constant returns to scale and use of the Armington assumption. SPE CGE models assume that goods within an industry are perfectly substitutable, and are generally employed to study spatial patterns of production. NEG CGE models relax the assumptions of perfect competition and constant returns to scale to account for agglomeration effects.

CGE models for congestion and externalities incorporate detailed household utility functions, particularly in disaggregating transport commodities, as well as additional constraints. These models represent interactions between agents beyond markets and the price mechanism, particularly through congestion, which also limits their ability to capture spatial detail.

5. DISCUSSION
The CGE models in the literature span a variety of scopes and interactions, both for freight and household transport. Many were developed to investigate projects requiring special analysis, for example to determine the impact to GDP from improved freight efficiency. While these models could be applied again to new projects that require it, from the perspective of planners and decision makers, the question is whether CGE models can go beyond the role of standalone assessment and be formalised as part of the appraisal process. Some jurisdictions are beginning to recommend the use of CGE models to derive inputs for appraisal in areas where the standard CBA process is considered lacking.

This section discusses how CGE models can contribute to the appraisal process by examining the capabilities of CGE models, how transport projects and policies can be assessed, how results can be used and under what circumstances. This section also discusses the theoretical and practical issues of implementing CGE models in appraisal and identifies areas for future research.

5.1 Transport appraisal using CGE models
To recap from earlier, the CBA is the most common form of appraisal for transport projects and policies. In a conventional CBA, the net social benefit of a proposal is calculated by combining valuations of impacts to users, service providers and the community, primarily through the consumer surplus of changes in the transport market. However, changes in transport will create impacts that spread throughout the wider economy, for example in land
and labour markets. The change in welfare from these impacts is theoretically captured in the consumer surplus of the transport market alone, even if their ultimate incidence is outside the transport market, if there are no technological externalities and markets operate perfectly. In reality, economies experience taxes, increasing returns to scale and other imperfections which prevent savings in transport costs passing between markets in full. These uncaptured impacts are known as WEIs, but their scale and even existence are controversial. WEIs are a new area of research and there is little consensus about how best to measure them. A further issue with CBAs is that they cannot identify the long-term distribution of impacts among markets and consumers without additional modelling.

CGE models simulate the behaviour of all markets in an economy linked together, and thus can trace the effects of a transport improvement flowing through to other markets. The responses of economic agents and markets iterate continuously until equilibrium, yielding the long-term distribution of impacts. This provides a holistic and flexible framework for estimating WEIs as the mechanisms of agglomeration can be incorporated from first principles, including linkages between markets and knowledge spillovers (but with consequences for the number of equilibria, as in NEG models), as well as other market imperfections. Nearly any economic metric can be extracted, such as welfare measures (e.g. equivalent variations and consumer surplus), GDP and other economic indicators. As model outputs will reflect both direct and indirect effects working in tandem, it can be difficult to disentangle the change in welfare attributable to WEIs. One method is to separately calculate the direct effects using conventional methods, and then subtract this from the total effects from the CGE model. This method was used in Hof et al. (2011) for CGEurope. In other models with detailed modelling of imperfect markets, individual surpluses from firm profits, taxes and other sources of lost welfare can be summed.

Currently, only the Netherlands recommends the use of CGE models to estimate WEIs in their transport appraisal guidelines (Wangsness et al., 2016). Instead of applying formulae to estimate welfare uplift as a proportion of direct impacts, this approach allows for both negative and positive WEIs according to market conditions. It has also enabled the Netherlands to monetise more WEIs than in any other guidelines since each market imperfection can be incorporated into the behaviour of the CGE model. These include the WEIs of agglomeration externalities, labour market effects and impacts in markets with imperfect competition, as well as impacts from inefficient land-use regulation. There are other appraisal guidelines which recognise CGE models, but only for use in a supplemental economic impact analysis (Transport for NSW (2013) for example). This view relates more to mainstream CGE modelling which focuses on macroeconomic results rather than welfare.

CGE models can also offer a different perspective for measuring the total change in welfare from a transport improvement. Household agents are simulated directly in a CGE model and transport demand can be generated from household activities. This adds significant detail to their behaviour, enabling changes in welfare to be measured from utility functions at the household level. For example, welfare can be calculated accounting for the preference of a consumer to spend a certain amount of time travelling per day as documented in empirical literature, whereas a CBA would assume all travel time savings are valued at a constant rate with an exogenous value of time. This is due to the convexity of household preferences, where the marginal utilities of leisure and income (through labour) increase as travel times increase. This was reflected in Anas (2015) using RELU–TRAN (Anas and Liu, 2007), where households substituted away from travel-intensive activities and locations as congestion increased, resulting in slower growth of travel times than population.

Metz’s (2008) concern with the conventional CBA approach was that the focus on travel time savings is misleading as they tend to evaporate when behaviour adjusts, and that it is accessibility that is actually valued. This perspective is accommodated in a CGE
framework since travel demand, as well as demand in all other markets, is elastic. Depending on the formulation, household utility in a CGE model will rise from a transport improvement due to increases in consumption and leisure, even if travel demand also rises as a result to negate the travel time savings that may be initially present. Other measures of welfare can also be extracted, such as the changes in consumer surplus across all markets.

Despite their advantages, it is impractical to supplant CBAs with CGE models for appraisal. Formulating, calibrating and running a CGE model takes considerable time, data and effort, and the level of spatial detail cannot match that of a CBA. For example, appraising a small road intersection upgrade would be straightforward with a CBA, but would hardly warrant the use of a CGE model as the effects would likely be negligible in the wider economy. The CGE results would also be highly spatially aggregated, and disaggregating the model would lengthen run times to an impractical extent. A CGE analysis is most worthwhile when the transport improvement is expected to impact the economy—determining when this applies requires experience. Another significant problem is the lack of transparency (‘black box’ nature) of CGE models—their operation tends to be difficult to understand, and practitioners may not be happy to take outputs as given (Vickerman, 2007b). Unfortunately, this is an inherent problem of CGE modelling as the solution process involves significant computations. Koopmans and Oosterhaven (2011) suggest one method of alleviating concerns is to run the model with and without certain behaviours active to estimate the contribution of each behaviour to the final outcome. Finally, appraisals usually require a time series of costs and benefits, which can only be delivered with a CGE model if it is dynamic.

For now, aside from generating WEIs, the best use of CGE models in appraisal may be to assess welfare from the household perspective, which can then be compared with welfare from conventional CBAs. This would provide an appraisal that would ensure no double counting, when welfare is assessed from the single point of measurement of household agent utility, and provide an analysis of nth order effects. A CGE appraisal would also enable policy makers to determine the distribution of impacts among markets and economic agents, and would facilitate comparisons with investments in other sectors of government spending. It may even be possible to integrate CGE models and conventional CBA methods to an extent, for example by deriving parameters for the CBA such as values of time from the CGE model.

Some of the models in this review are more suitable for appraisal than others. RELU–TRAN (Anas and Liu, 2007) would be appropriate for an urban infrastructure project as transport is simulated as a network. RAEM, on the other hand, lacks the integrated network representation but introduces detail in modelling agglomeration economies and macroeconomic behaviour. Partridge and Rickman (2010) suggested a number of features that regional CGE models should incorporate to be useful for policy analysis. The location behaviour of agents should be influenced by the attractiveness of a region, whether by allowing for regionally-differentiated taxes in the model, or by adding a factor to account for the consumption of local amenities. Urban models such as RELU–TRAN do include parameters for the inherent attractiveness of regions (Anas and Liu, 2007), but being parameters, they are not a function of model variables such as regional government expenditure. Features from NEG such as mechanisms for agglomeration externalities are necessary in some circumstances. Models should also have well-specified linkages to account for openness between regions, with friction. This would particularly apply to labour markets, where the assumptions of perfect mobility and full employment should be relaxed, and labour forces should be differentiated by skill. Tavasszy et al. (2011) provided guidance from experience in the Netherlands that modellers should pay close attention to the specification of transport costs and sources of irrational agglomeration in model behaviour.
The models examined in this review provide the ingredients to construct a full model for transport appraisal. Firstly, models from urban economics provide the theory to account for land markets and migration, and demonstrate how discrete choice models can be integrated. Secondly, spatial CGE models can describe the behaviour of macroeconomic agents. Thirdly, models from the SPE literature describe methods of linking transport network models and CGE models. Fourthly, NEG models account for the behaviour of regional markets, particularly in terms of imperfect markets, and resulting patterns of spatial development. Finally, models from congestion and externality modelling have detailed representations of household demand for transport and provide methods to account for community impacts, such as pollution and government spending.

The models in this review also demonstrate the wide variety of approaches for linking partial equilibrium transport models with CGE models. RELU–TRAN (Anas and Liu, 2007) and Rutherford and van Nieuwkoop’s (2011) model show two different methods for integrating traffic assignment models. In RELU–TRAN, the CGE and traffic assignment models are kept independent. Each iteration of the full model involves individually solving the CGE and traffic assignment models, with the outputs of one model feeding into the other, until both models have converged together. This process of sequentially updating independent models was adopted in many other cases such as MITER (Roson, 1996). On the other hand, in Rutherford and van Nieuwkoop’s model, the CGE and traffic assignment equations are combined into a single mathematical problem. Other types of partial equilibrium transport models have also been linked with CGE models. In TRESIS–SGEM (Hensher et al., 2012; Truong and Hensher, 2012), departure time and mode choice are estimated with a nested discrete choice structure conditional on CGE outputs, which eventually feed back into prices in the CGE model. Another example is the USAGE-Hwy model (Dixon et al., 2017), which is a conventional CGE model that has been designed to take standard transport network model inputs, including travel times and costs, as well as more specific transport model inputs such as fuel use by trucks and cars, fatalities and vehicle maintenance costs. However, in USAGE-Hwy, the CGE outputs are not fed back into the transport model.

Constructing a model with all of these aspects is a formidable task as there would be obstacles in finding a solution (there would be multiple equilibria) and interpreting the solution, as well as data and computational requirements. Nevertheless, this review should provide a menu of the wide variety of behaviours and applications that have already been studied in CGE models for transport, such that future modellers can understand how they have been accomplished and incorporate the aspects relevant to their project.

5.2 Theoretical and practical issues of CGE modelling

The review to this point has highlighted where CGE modelling can complement the transport appraisal process, but more general methodological concerns remain about their implementation in appraisal.

In light of the history of CGE modelling described in Section 3.1, general equilibrium theory is often cited as the theoretical foundation of all CGE models, both those from the IO branch and the GE theory branch. The theory, based on the Arrow–Debreu (1954) model, was considered to be a crowning achievement of 20th century neoclassical economics and is still central to economic practice. Much of its esteem derives from two fundamental results: (1) that any market equilibrium of an economy consisting of perfectly competitive agents and without market failures is a Pareto optimum, and (2) any Pareto optimum can be achieved with lump-sum transfers. According to Ackerman (2002), these results are often employed in a normative context to assert properties about the efficiency and equity of market economies, and to claim that any preferred equilibrium can be sustained with appropriate transfers.
However, general equilibrium theory has faced criticisms from some quarters. Notwithstanding questions about whether central assumptions such as constant returns to scale and perfect competition are realistic, the Sonnenschein–Mantel–Debreu theorem proves that aggregate excess demand functions inherit few of the properties of agent excess demand functions—mainly continuity and homogeneity of degree zero in prices. Many of the other assumptions of agent rational behaviour (e.g. convexity) can be lost when agents are aggregated. This is due to agents differing in their reaction to an increase in the price of a good—demand from some agents might lower due to the substitution effect, but rise from other agents due to the wealth effect. In aggregate, the effect that dominates might switch as prices change, such that the aggregate agent does not behave ‘rationally’ and equilibrium is achieved at multiple sets of prices. This was problematic for general equilibrium theory and other fields relying on aggregate demand functions as comparative static analysis may be difficult to interpret when there is no corresponding equilibrium between scenarios. In addition, the theorem leads to results demonstrating that economies may not converge to Arrow–Debreu equilibria and that they may not be stable (Ackerman, 2002).

While these are concerns regarding some of the original claims of general equilibrium theory, the question is to what extent they translate to CGE models and whether they impact their validity. Firstly, CGE models are variable in their closeness to the theoretical model. While some CGE models from the GE theory branch are a pure application of the theoretical model (as described in Section 3.1), others from the IO branch are further removed as the ratios of demand in the national income identity are fixed. Regardless, the Sonnenschein–Mantel–Debreu theorem does not disprove their operation.

If CGE models are taken as an empirical framework to model the economy in terms of prices, quantities and agents, separate from the normative aspects of general equilibrium theory, they will have value if they can accurately predict the state of the economy and can be modified to incorporate more realistic behaviour as theory becomes available. This is a similar argument to those that justify static traffic models—although traffic models may not be ‘correct’ in replicating human decision-making processes, they provide reasonable predictions about the long-term behaviour of the system. In the case of CGE models, the tendency for markets to equilibrate over time is not contentious—how equilibrium is achieved is what is debated. Ackerman (2002) believes that social and institutional constraints stabilise the market system. CGE models do not replicate these dynamics, but instead only predict the eventual position of the economy.

The principle behind their use in appraisal is that even though economies comprise individual transactions, prices will tend towards equilibrium over time, which provides a consistent basis for measuring impacts and comparing proposals. CGE models seem particularly well suited to assessing infrastructure as the simulated ‘shocks’ have a physical interpretation—they represent the change in infrastructure, which is an exogenous factor that drives the model. In any case, the behaviours underlying CGE models are based on the same economic theory underlying conventional appraisals, including the basic optimisation models used to derive the value of time. CGE models are in essence an extension of conventional appraisals through the introduction of a price mechanism to simulate movements in all markets.

Some of the concerns about general equilibrium theory, such as multiple equilibria, can be considered realistic features in CGE models. In the NEG literature, multiple equilibria are a natural and expected outcome of agglomeration effects, and the movement from a prior equilibrium to another is analysed in terms of dynamic ‘forces’. On the other hand, most fields of CGE modelling either use functional forms that result in unique equilibria or assume that a simulation will converge to a nearby equilibrium. One reason for this may be that simulations tend not to stray far from established equilibria, and Bröcker and Mercenier
(2011) state that no studies using calibrated CGE models have reported multiple solutions. There has been little analysis of how static models converge to a particular equilibrium and whether the choice of solver has an influence. This is related to a general need to study the dynamics of how economies converge and under what conditions.

Furthermore, the fundamental problem posed by the Sonnenschein–Mantel–Debreu theorem, that macroeconomic behaviour can only be predicted from detailed information about individual behaviour (Geanakoplos, 1987), is still relevant to CGE modelling as there is a heavy reliance on representative agents. Individual agents can only be aggregated into a single, well-behaved representative agent if their utility functions are restricted to the Gorman polar form (Gorman, 1961).

One solution in recent literature has been to link microsimulation models of economic agents with aggregate CGE models (Bourguignon et al., 2005). In this approach, economy-wide variables are calculated at the CGE level, which then become exogenous parameters for the microsimulation level. While this enables within-group variability to be captured, it can be computationally expensive to iterate between the stages and allow microsimulation agents to interact. Magnani and Mercenier (2009) described an intermediate approach where representative agents are disaggregated according to standard categorisations used in discrete choice models. They demonstrated that aggregates of uniform individuals modelled using discrete choice will behave identically to agents in CGE models when represented with constant elasticity of substitution/transformation functions. This result, based on the equivalence between discrete choice and continuous demand functions described in Section 4.1 (Anderson et al., 1989), provides justification for the typical assumed behaviour of representative agents in CGE models. Whether this disaggregation could impact the stability of equilibria requires further investigation.

There are still many aspects of the economy that are difficult to represent in a CGE framework, for example demand for money or share trading. The ability to model large structural changes is questionable as elasticities of substitution are normally constant and production functions assume fixed input ratios. However, many of Ackerman’s (2002) suggestions for rectifying general equilibrium theory can be incorporated in a CGE framework. Models of consumer behaviour can be formulated to account for bounded rationality and mechanisms for simulating social interaction have been developed to an extent in CGE models of externalities. Ackerman noted that simulations of the latter could potentially cause feedback loops, but such effects are already familiar in CGE models of agglomeration.

Barriers also remain to the practical implementation of CGE models. Calibrating a CGE model requires the specification of a ‘benchmark’ dataset representing transactions between all agents in an economy, typically in the form of an input–output table or social accounting matrix. This dataset can be difficult to obtain or expensive to create, more so if it is spatially disaggregated. As CGE models are calibrated to replicate the benchmark dataset when no shocks are applied, there is an assumption that the benchmark dataset represents an economy at equilibrium. Some models are calibrated to time-series data, but this is the exception rather than the norm due to the substantial data requirements. Giesecke (2002), for example, allowed for the dynamic variation of parameters in the FEDERAL-F model. Statistical estimation of parameters is difficult due to the large number of observations required as well as their partitioning into price and quantity variables (Shoven and Whalley, 1992). Miyagi (1998) and Ando and Meng (2009) claim that the calibration methods of CGE modelling are actually an advantage as they are less data intensive than comparable econometric methods since large data samples are not required for regression formulae. A notable problem is the calibration of elasticities of substitution, for which there are limited
and sometimes contradictory studies available to inform their estimation. As a result, they are often set using the personal judgement of the modeller.

Validation of CGE models is another area of concern to modellers. Even though the very first CGE model by Johansen (1960) was validated over a 10 year dataset, it is rare to see the forecasts of CGE models tested with external time-series data. CGE models will replicate the benchmark dataset perfectly due to the calibration process. From here, the best that is usually done is to test the consistency of the model to check for errors in coding and data-handling. Sensitivity analysis may also be performed to test the robustness of the model to errors in parameter estimation, and results from simulations are often compared with back-of-the-envelope models\(^7\) to check that they can be explained intuitively.

For modellers who are only interested in the qualitative effects of economic changes, this level of validation may be adequate, but others may find it unsatisfactory for the precise calculations required in an appraisal. Dixon and Rimmer (2013) suggested that a model can be tested by its ability to replicate historical data. Kehoe et al. (1995) analysed the performance of a CGE model of Spain, 10 years after its estimation, finding that its results were generally accurate. Kehoe (2003) later found that CGE models applied to NAFTA performed poorly, emphasising the need for ex-post evaluations of models to inform future models and improve confidence in the field. Partridge and Rickman (2010) advocated for time-series calibration with historical data validation to become standard practice.

Models are not only sensitive to parameters, but also the functional forms chosen by the modeller (Sundberg, 2005). McKitrick (1998) demonstrated that even with the same data, different functional forms can have a significant influence on model results, which highlights the importance of justifying why a particular structure is chosen and validating its performance. In general, validation continues to be acknowledged as an area of CGE modelling requiring development.

6. CONCLUSION
Technological advances have made the application of sophisticated simulation models of economic and transport systems viable in recent decades. This has provided the opportunity to improve on past methods of transport appraisal by relaxing assumptions inherent in the static formulae used to estimate impacts. One type of model that has become increasingly popular to analyse transport projects and policies is the CGE model. These models simulate every market in an economy through the actions of consumers, producers and other economic agents. Being built from microeconomic first principles, it is possible to extract a rich array of outputs and represent relationships between transport and the economy that would be difficult to simulate in any other model. However, CGE models are still unknown to many in the transport planning domain, and there has not yet been a full review of CGE models from the transport planning perspective. Therefore, this paper aimed to provide such a review, and to discuss their potential application in the appraisal process.

On the first aim, this review examined models from a range of spatial scales, beginning with urban CGE models. These models incorporate discrete choice structures, assume free movement of residences and employment, and simulate land markets. As such, when linked with transport network models, they are suitable for urban infrastructure appraisal. Regional CGE models tend to be freight-oriented and have emerged from three strands of literature: spatial CGE modelling, SPE modelling and NEG modelling. Spatial CGE models were developed by extending conventional single-region models into the spatial

---

\(^7\) A ‘back-of-the-envelope’ model is a simplified model of the full model, which modellers often use to explain more complex interactions within the full model.
dimension. They normally assume that regional varieties of a commodity are imperfect substitutes, and are more likely to include macroeconomic agents such as the government, investors and foreign trade. SPE-based CGE models instead assume, to an extent, that regional varieties are perfect substitutes. They are also more likely to be combined with transport models. NEG-based CGE models account for imperfect markets to explain agglomeration and spatial development. Finally, there are non-spatial CGE models that account for externalities such as congestion and pollution through detailed representations of household behaviour. These models are used for policy analysis as they can simulate the effects of congestion pricing, emissions pricing and others.

On the second aim, this paper explained the two well-known shortcomings of conventional CBAs, the most common type of appraisal. Firstly, metrics from conventional CBAs do not account for imperfect markets and externalities, and secondly they do not provide the long-term distribution of benefits. Both of these are highly relevant to transport planning nowadays. For the former, transport projects are often justified in part by their potential for economic development, and a range of ad-hoc methods have been used to estimate WEIs in recent appraisal practice. For the latter, knowing the distribution of benefits is both socially and politically important in project prioritisation.

CGE models can provide a unified framework to estimate WEIs as well as the distribution of both direct and indirect impacts of transport improvements. In a CGE model, nearly any linkage between transport and the economy can be simulated, and parameters that would be static in a CBA can be made endogenous. Utility can be measured at the household level after responses of economic agents and markets have iterated until equilibrium. Metrics such as equivalent variations, consumer surplus and GDP can then be extracted. However, CGE models as a method of appraisal cannot replace CBAs as they are costly to build and are more spatially aggregated. At this stage, it may be most appropriate to use CGE models to extend conventional CBAs, integrate their outputs with CBAs (e.g. to estimate parameters for CBAs) or to use them as an alternative method of appraisal for comparison. There are also a number of issues regarding calibration and validation that may need to be resolved before CGE models are acceptable to transport planning practice.

For practitioners, this review can be taken as a record of the variety of CGE models applied to transport, so that a suitable model can be selected to simulate the impacts of a proposed project or policy. The review can also be used to identify features that might be desirable in developing a new model, or to understand how a particular model fits in the broader world of CGE modelling. From the discussion in this paper, it is hoped that the adoption of CGE modelling as a transport planning tool can improve the quality of transport appraisals and the decisions based on them to promote the equitable and rational development of transport networks worldwide.

7. ACKNOWLEDGEMENTS
This work was supported by an Australian Government Research Training Program Scholarship.

The authors would like to thank three anonymous reviewers whose comments have greatly improved the paper.

8. REFERENCES
Ackerman, F., 2002. Still dead after all these years: interpreting the failure of general equilibrium theory. Journal of Economic Methodology 9, 119–139.
https://doi.org/10.1080/13501780210137083


North-West Rail Link project in Sydney, Australia. Journal of Transport Geography 24, 292–305. https://doi.org/10.1016/j.jtrangeo.2012.03.009


Sundberg, M., 2005. Spatial Computable General Equilibrium Modeling - Static and Dynamic Approaches. Royal Institute of Technology.


APPENDIX A: LIST OF ABBREVIATIONS

Table A.1 lists the abbreviations used in this paper. Model and software names have not been included as they are generally known by their abbreviation rather than their full name, if it exists.
Table 1: List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>applied general equilibrium</td>
</tr>
<tr>
<td>AMM</td>
<td>Alonso–Mills–Muth</td>
</tr>
<tr>
<td>BCR</td>
<td>benefit–cost ratio</td>
</tr>
<tr>
<td>CBA</td>
<td>cost–benefit analysis</td>
</tr>
<tr>
<td>CGE</td>
<td>computable general equilibrium</td>
</tr>
<tr>
<td>CIF</td>
<td>cost, insurance and freight</td>
</tr>
<tr>
<td>CUE</td>
<td>computable urban equilibrium</td>
</tr>
<tr>
<td>FOB</td>
<td>free on board</td>
</tr>
<tr>
<td>GDP</td>
<td>gross domestic product</td>
</tr>
<tr>
<td>GE</td>
<td>general equilibrium</td>
</tr>
<tr>
<td>GTAP</td>
<td>Global Trade Analysis Project</td>
</tr>
<tr>
<td>IO</td>
<td>input–output</td>
</tr>
<tr>
<td>LUTI</td>
<td>land-use transport interaction</td>
</tr>
<tr>
<td>MARKAL</td>
<td>market allocation</td>
</tr>
<tr>
<td>MCA</td>
<td>multi–criteria analysis</td>
</tr>
<tr>
<td>NAFTA</td>
<td>North American Free Trade Agreement</td>
</tr>
<tr>
<td>NEG</td>
<td>new economic geography</td>
</tr>
<tr>
<td>OD</td>
<td>origin–destination</td>
</tr>
<tr>
<td>RUBMRIO</td>
<td>random-utility-based multiregional input–output model</td>
</tr>
<tr>
<td>SACTRA</td>
<td>Standing Advisory Committee for Trunk Road Assessment</td>
</tr>
<tr>
<td>SCGE</td>
<td>spatial computable general equilibrium</td>
</tr>
<tr>
<td>SPE</td>
<td>spatial price equilibrium</td>
</tr>
<tr>
<td>WEI</td>
<td>wider economic impact</td>
</tr>
</tbody>
</table>